



Comparative full-length transcriptome analysis by Oxford Nanopore Technologies reveals genes involved in anthocyanin accumulation in storage roots of sweet potatoes (*Ipomoea batatas* L.)

Jun Xiong^{1,2}, Xiuhua Tang², Minzheng Wei² and Wenjin Yu¹

¹Agricultural College, Guangxi University, Nanning, China

²Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China

ABSTRACT

Background. Storage roots of sweet potatoes (*Ipomoea batatas* L.) with different colors vary in anthocyanin content, indicating different economically agronomic trait. As the newest DNA/RNA sequencing technology, Oxford Nanopore Technologies (ONT) have been applied in rapid transcriptome sequencing for investigation of genes related to nutrient metabolism. At present, few reports concern full-length transcriptome analysis based on ONT for study on the molecular mechanism of anthocyanin accumulation leading to color change of tuberous roots of sweet potato cultivars.

Results. The storage roots of purple-fleshed sweet potato (PFSP) and white-fleshed sweet potato (WFSP) at different developmental stages were subjected to anthocyanin content comparison by UV-visible spectroscopy as well as transcriptome analysis at ONT MinION platform. UV-visible spectrophotometric measurements demonstrated the anthocyanin content of PFSP was much higher than that of WFSP. ONT RNA-Seq results showed each sample generated average 2.75 GB clean data with Full-Length Percentage (FL%) over 70% and the length of N50 ranged from 1,192 to 1,395 bp, indicating reliable data for transcriptome analysis. Subsequent analysis illustrated intron retention was the most prominent splicing event present in the resulting transcripts. As compared PFSP with WFSP at the relative developmental stages with the highest (PH vs. WH) and the lowest (PL vs. WL) anthocyanin content, 282 and 216 genes were up-regulated and two and 11 genes were down-regulated respectively. The differential expression genes involved in flavonoid biosynthesis pathway include *CCoAOMT*, *PpLDOX*, *DFR*, Cytochrome P450, *CHI*, and *CHS*. The genes encoding oxygenase superfamily were significantly up-regulated when compared PFSP with WFSP at the relative developmental stages.

Conclusions. Comparative full-length transcriptome analysis based on ONT serves as an effective approach to detect the differences in anthocyanin accumulation in the storage roots of different sweet potato cultivars at transcript level, with noting that some key genes can now be closely related to flavonoids biosynthesis. This study helps to improve understanding of molecular mechanism for anthocyanin accumulation

Submitted 19 October 2021

Accepted 16 June 2022

Published 12 July 2022

Corresponding authors

Minzheng Wei,

Weimzheng2021@126.com

Wenjin Yu, Yuwjn163@163.com

Academic editor

Atsushi Fukushima

Additional Information and
Declarations can be found on
page 12

DOI 10.7717/peerj.13688

© Copyright

2022 Xiong et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

in sweet potatoes and also provides a theoretical basis for high-quality sweet potato breeding.

Subjects Bioinformatics, Molecular Biology, Plant Science

Keywords Sweet potato, Oxford Nanopore Technologies, Full-length transcriptome, Anthocyanin, Storage roots

INTRODUCTION

Sweet potato (*Ipomoea batatas* L.) is widely planted worldwide for its high output, available management, high nutritional value and wide range of uses (Andreason & Olaniyi, 2021). With respect to sweet potato production, storage roots constitute the most economically important agronomic trait, which play an important role in nutrient storage and reproduction. Sweet potato cultivars distribute globally and their storage roots show different skin color (such as white, yellow, cream, red, pink and orange) and flesh color (including orange, white, yellow, cream and purple) (Wang et al., 2018). Sweet potato is abundant in β -carotene, carbohydrates, minerals and dietary fiber (De Moura, Miloff & Boy, 2015; Sun et al., 2014), nutrient variation of which depends to a large extent on cultivar varieties. The nutrient composition of several cultivars has been studied, including the white-fleshed sweet potato (WFSP), purple-fleshed sweet potato (PFSP) and yellow- or orange-fleshed sweet potato cultivars. The amount of ash, protein, crude fiber, total reducing sugars, and β -carotene exhibited significant difference between yellow and white sweet potato cultivars which were planted in Rwanda (Wang et al., 2018). Antioxidant activities of storage roots varied widely among the sweet potato cultivars, and natural colorants such as anthocyanin that is rich in purple and red-fleshed sweet potatoes inclined to related to high antioxidant activity (Ghasemzadeh et al., 2016; Sun et al., 2019). The high antioxidant effects demonstrated that red and purple fleshy sweet potatoes have potential use in the nutraceutical industry (Wang et al., 2020). Flavonoids including anthocyanin have been found in many plants and are conducive to plant environmental adaptation (Kovinich et al., 2014), fruit development (Jaakola, 2013; Petroni & Tonelli, 2011), and even human health (Alipour, Rashidkhani & Edalati, 2016). They accumulate in various tissues of plants at different developmental stages (Hichri et al., 2011), causing color variation in plant tissues such as fruits of kiwifruit (Man et al., 2015) and storage roots of sweet potatoes (Li et al., 2021). Flavonoid biosynthesis is regulated by a diverse array of exogenous and endogenous factors present in secondary metabolic pathways, which is widely researched in many plants. However, the molecular mechanism of flavonoid biosynthesis in sweet potato is not fully elucidated and few studies attempted to reveal genes related to anthocyanin accumulation in storage roots of sweet potatoes.

RNA sequencing (RNA-Seq) has been an effective method (Uapinyoying & Goecks, 2020) to explore novel genes and was wildly employed to study the molecular mechanisms for growth and development in plants (Younesi-Melerdi, Nematzadeh & Pakdin-Parizi, 2020). Oxford Nanopore Technologies (ONT) MinION as the newest third-generation sequencing technology (TGS), has obvious advantages in achieving the specific sequencing

of full-length transcripts, quantification of isoform level, and identification of complex gene structures such as alternative splicing, polyadenylation, and fusion genes (Slatko, Gardner & Ausubel, 2018). In this research, the transcriptome profiles of storage roots of PFSP and WFSP at different developmental stages were identified using the ONT MinION platform in order to find genes involved in anthocyanin accumulation in storage roots of sweet potatoes. The attempt is beneficial to improving molecular understanding of anthocyanin accumulation in sweet potatoes and also provides a theoretical basis for high-quality sweet potato breeding.

MATERIALS AND METHODS

Plant materials

Purple-fleshed sweet potato (PFSP: Guijingshu 8) and white-fleshed sweet potato (WFSP: the mutant strain 118 of Guijingshu 8) were bred by institute of Cash Crops, Guangxi Academy of Agricultural Sciences. Both cultivars were planted in the base of Guangxi Academy of Agricultural Sciences in August (Latitude: 22.853076, Longitude: 108.245977 Altitude: 55.59). The two cultivars were planted in five rows, the length was 5 m, the row spacing was 1.0 m 0.2 m, and 14 plants were planted in each row.

Samplings were conducted in 40, 50, 60, 70, 85, 95 and 110 days after planting. Three tubers from each cultivars in every stages were taken and washed with clean water. Cultivars in every stages were taken photos and arranged in two rows in Fig. S1 (PFSP in the first row and WFSP in the second row). Each tuber was divided into two parts. One was quickly sliced and put into a 2 ml centrifuge tube with forceps and frozen in liquid nitrogen for use in full-length transcriptome. The other was cut into 1 cm × 1 cm × 1 cm tubers, dried at 40 °C for 72 h, and was used for determination of anthocyanin content. A total of 1 g sweet potato powder was extracted with 0.5 mol/L citric acid buffer (pH 3) in 60 °C water bath for 2 h. The extracted liquid was cooled, 1 ml supernatant was taken and centrifuged at 12,000 rpm for 5 min. The light absorption value of the supernatant was measured at 530 nm. The detail was in related literature (Hong, Netzel & O'Hare, 2020).

After sampling in the whole growth period is completed, the anthocyanin content in each period is measured to understand the change trend of anthocyanin content. Samples from two periods of low and high anthocyanin content in two cultivars were taken for RNA sequencing. The corresponding groups are listed in the Table S1.

RNA extraction, library construction, and sequencing

Total RNA was extracted from 12 samples which were divided into four groups (three replicate samples for each group). The NEBNext Poly(A) mRNA Magnetic Isolation Module was employed to enrich poly(A) mRNA to extract total RNAs from 12 root samples. We synthesize the cDNA for sequencing according to the strand-switching protocol from Oxford Nanopore Technologies. Firstly, full-length cDNA libraries were prepared from the poly(A) mRNAs by using the cDNA-PCR Sequencing kit by Oxford Nanopore (SQKPCS109) (Yao et al., 2020). Then, PCR was employed to amplify the cDNA for 13 to 14 cycles with specific barcoded adapters from the Oxford Nanopore PCR Barcoding kit (SQKPBK004). Finally, before loading onto a PromethION flow cell

in a MinION sequencer, the 1D sequencing adapter was synthesized to the DNA. The sequencing was running on MinKNOW. The sequencing data were loaded to the Sequence ReadArchive (SRA), National Center for Biotechnology Information (NCBI), and the accession number is [PRJNA717378](#).

Transcriptome analysis, lncRNA, differentially expressed transcript (DET) identification and functional annotation

Firstly, raw reads were screened by following standard: minimum average read quality score of 7 and minimum read length of 500bp. After mapping to the rRNA database, we discarded the reads corresponding to the rRNA. Next, we identified the full-length, non-chemiric (FLNC) transcripts by searching for primer at both ends of the reads. After minimap2 mapping to the reference genome, FLNC transcripts clusters were obtained, and the consensus isoforms were screened after polishing within pinfish in each cluster ([Hu et al., 2020](#)).

Subsequently, the minimap2 was used to map the consensus sequences to reference genome ([Hoang et al., 2018](#)). The cDNA_Cupcake package was utilized to collapse the mapped reads with min-coverage of 85% and min-identity of 90%. When folding the redundant transcripts, the 5' difference were not considered.

Power analysis of the transcriptome data was done using the RNASeqPower package. The statistical power among different experimental groups were 0.84 (PH vs. WH) and 0.73 (PL vs. WL) respectively.

Four calculation methods of CPC, CNCI, CPAT and Pfam were combined to sort non-protein coding RNA candidates from putative protein-coding RNAs in the transcripts. The minimum length and exon number threshold filtered out the putative protein-coding RNAs. LncRNA candidates were screened from transcripts according to the standard: a length of more than 200 nt and more than two exons. Then, CPC/CNCI/CPAT/Pfam were used to distinguish the protein-coding genes from the non-coding genes for further screening ([Zhang et al., 2019](#)).

The DESeq R package (1.10.1) was utilized to analyze the differential expression analysis of two groups. DESeq provides a statistical procedures model based on the negative binomial distribution to detect the differential expression in digital gene expression data. The Benjamini and Hochberg's approach were used to adjust the resulting *P* values ([Benjamini & Hochberg, 1995](#)) to control the false discovery rate. The differentially expressed genes were assigned with a FDR < 0.01 and foldchange hang found by DESeq.

Gene function was annotated by using the following databases: GO (Gene Ontology), NR (NCBI non-redundant protein sequences), KEGG (Kyoto Encyclopedia of Genes and Genomes), Pfam (Protein family), KOG/COG/eggNOG (Clusters of Orthologous Groups of proteins) and Swiss-Prot (A manually annotated and reviewed protein sequence database). Statistical analyses were performed using the SPSS 22.0 software package (IBM SPSS, Somers, NY, USA) ([Zhao et al., 2019](#)).

Six genes in flavonoid biosynthesis pathway were screened for validation using quantitative real-time PCR (qRT-PCR). The primers used for qRT-PCR were listed in [Table S7](#). The LightCycler®480 II Real-Time System (LightCycler®480 II cycler, Roche,

Carlsbad, CA, USA) was utilized to perform the qRT-PCR in a 96-well plate. The thermal profile for the PCR amplification was 95 °C for 5 min, followed by 40 cycles of 10 s at 95 °C and another 40 cycles at 60 °C for 30 s. According to the instruction's protocol, all the PCR reactions were performed using the Hieff™ qPCR SYBR Green Master Mix (No Rox) (Yeasen Biotech Co., Ltd., Shanghai, China). All the qRT-PCR analyses were conducted with three technical and three biological replicates. According to the $2^{-\Delta\Delta CT}$ method, we calculated the expression level of different genes to the control (*Livak & Schmittgen, 2001*).

Data processing and analysis

The SPSS 22.0 software package (IBM SPSS, Somers, NY, USA) was utilized to perform the statistical analyses.

RESULTS

Anthocyanin content variation

In this work, two widely grown sweet potato cultivars in Guangxi province (China) viz. PFSP with purple-colored skin and WFSP with red-colored skin were studied for anthocyanin content variation in storage roots. Visual inspection of the sweet potato cultivars showed that PFSP exhibited more purple pigments than WFSP (*Fig. 1A*). Anthocyanin content in PFSP roots showed a trend of increase as planting days increased, with the highest level of anthocyanin detected at 95 days (*Fig. 1B*).

Regarding anthocyanin content in WFSP roots, at first it varied little during 40-70 days, then reached to the maximum at 85 days, and finally decrease from 85 to 110 days (*Fig. 1C*). As suggested, the anthocyanin content in PFSP roots was evidently much higher than that of WFSP (*Fig. 1*).

Full-length transcriptome sequencing analysis

ONT RNA-Seq approach (*Fig. 2A*) was applied for full-length transcriptome sequencing of 12 samples (PL, PH, WL and WH with three biological replicates), and subsequent transcriptomic analysis was conducted. After stringent quality checks and data cleaning, the total clean data was obtained, and the clean data of each sample reached 2.75 GB. The length of N50 ranged from 1,191 to 1,395 (*Table 1*).

According to the principle of cDNA sequencing, reads with primers identified at both ends are classified to be full-length sequences. The statistics of the full-length sequences obtained in this study could refer to *Table S2*. The minimum and maximum full-length percentage (FL %) of samples were 70.54% and 79.26%, respectively (*Table S2*). Lengths of multiple full-length reads were about 1.5 kb (*Fig. 2B*). The obtained full-length sequences were polished to acquire consensus isoforms. All the consensus transcripts (*Table S3*) were mapped to the reference genome through the minimap2 software for de-redundancy analysis, generating 33,709 mapped transcripts.

Alternative splicing analysis and functional annotation of the full-length transcriptome

Precursor mRNA produced in gene transcription can be spliced in many ways. Through alternative splicing, different exons are optioned to produce corresponding mature RNAs,

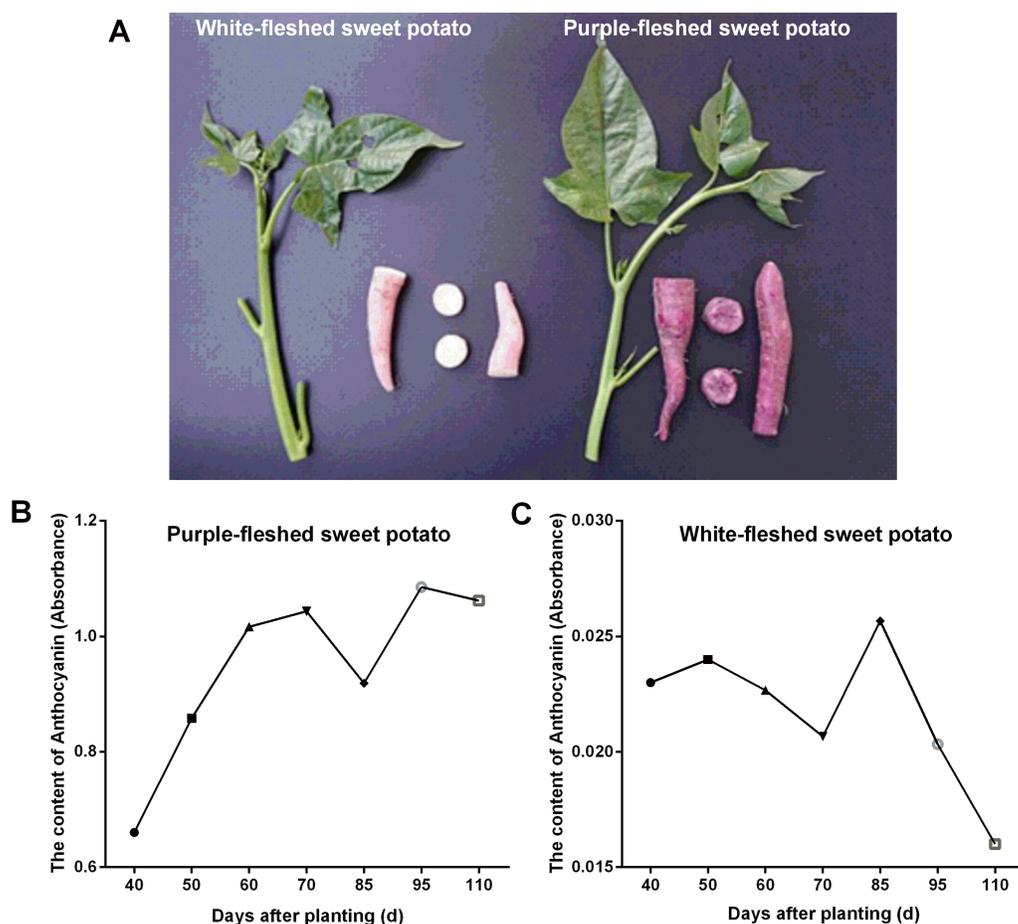


Figure 1 The phenotypes of sweet potatoes and anthocyanin content variation in their storage roots. (A) The phenotypes of PFSP and WFSP. (B) The anthocyanin profile of storage roots of PFSP during different periods. (C) The anthocyanin profile of storage roots of WFSP during different periods.

Full-size DOI: [10.7717/peerj.13688/fig-1](https://doi.org/10.7717/peerj.13688/fig-1)

which can be translated into relative proteins representing the diversity of biological traits (Lin & Krainer, 2019). Based on the analysis results of Astalavista software, the full-length transcripts were statistically divided into five alternative splicing events (Fig. 2C). Figure S2 exhibited the corresponding proportion of the five alternative splicing events for each sample. Among these five alternative splicing events, intron retention events (39.8%, 11,798) were the most prominent type of alternative splicing, followed by alternative 3' splice site (25.0%, 8427), alternative 5' splice site (16.0%, 5393), exon skipping (14.1%, 4961) and mutually exclusive exon (5.1%, 2500), which was consistent with the findings of previous reports on other plant species (Li et al., 2017; Wang et al., 2017a).

Meanwhile, a total of 4,045 transcripts were newly found and functionally annotated by the following databases with the number of mapped reads present in the brackets: NCBI non-redundant protein sequences (NR) (4031), Gene Ontology (GO) (1842), eukaryotic Ortholog Group (KOG) (1721), and Swiss-Prot Protein Sequence (Swiss-Prot) (2483) (Table S4).

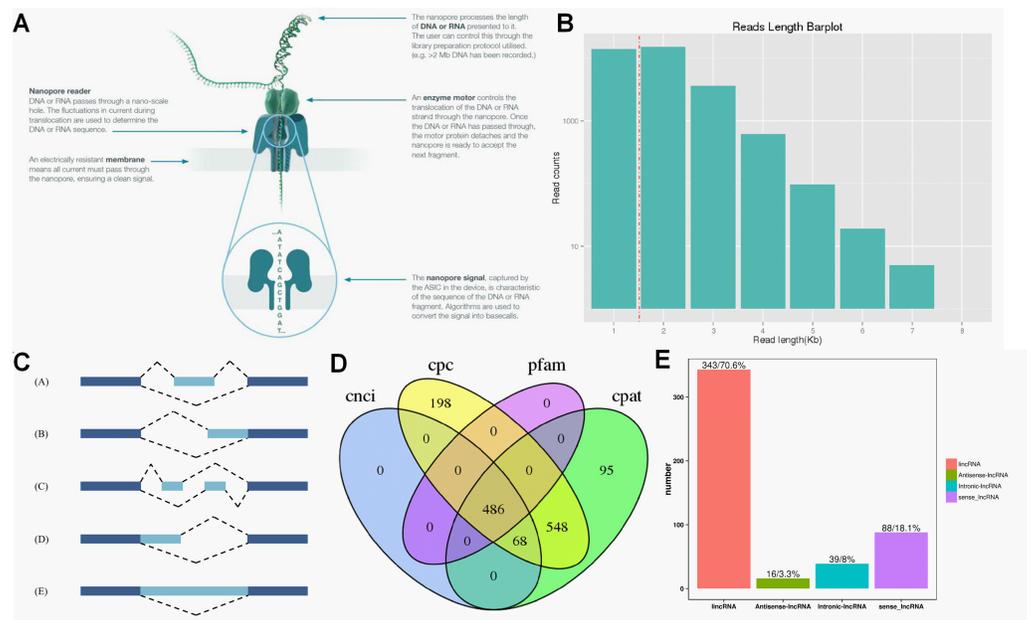


Figure 2 Overview of transcriptome data obtained by ONT. (A) The principle of ONT RNA-Seq approach. (B) Length of reads. (C) The five types of alternative splicing. (D) Classification of lncRNA. (E) The number of lncRNA.

Full-size DOI: 10.7717/peerj.13688/fig-2

Table 1 Full-length transcriptome data set.

#FileName	ReadNum	BaseNum	N50	MeanLength	MaxLength	MeanQscore
PH01	3447625	4.11E+09	1315	1191	11804	Q11
PH02	2592205	3.08E+09	1277	1186	14447	Q10
PH03	3386960	4.18E+09	1376	1233	12695	Q11
PL01	2324569	2.79E+09	1330	1201	13672	Q10
PL02	3173987	3.8E+09	1329	1196	11916	Q10
PL03	2254098	2.75E+09	1379	1220	14229	Q10
WH01	3380641	3.78E+09	1191	1118	9819	Q10
WH02	2460464	2.88E+09	1243	1170	12179	Q10
WH03	2471832	2.96E+09	1395	1198	11134	Q10
WL01	3295322	3.93E+09	1316	1192	10604	Q11
WL02	2288800	2.77E+09	1340	1210	11284	Q10
WL03	3013487	3.45E+09	1192	1143	14358	Q11

lncRNA prediction

Regulatory functions are one of the main functions of lncRNAs which are vital for post-transcription, transcription, and epigenetics (Huang *et al.*, 2018). CPC analysis, CNCI analysis, CPAT, and Pfam protein domain analysis were used to predict lncRNAs from transcriptome data. A total of 486 lncRNAs were predicted by CPC, CNCI, CPAT, and Pfam software (Fig. 2D). Subsequently, lncRNAs were classified and mapped according to the reference genome annotation information. The results indicated 343 (70.6%) lncRNAs

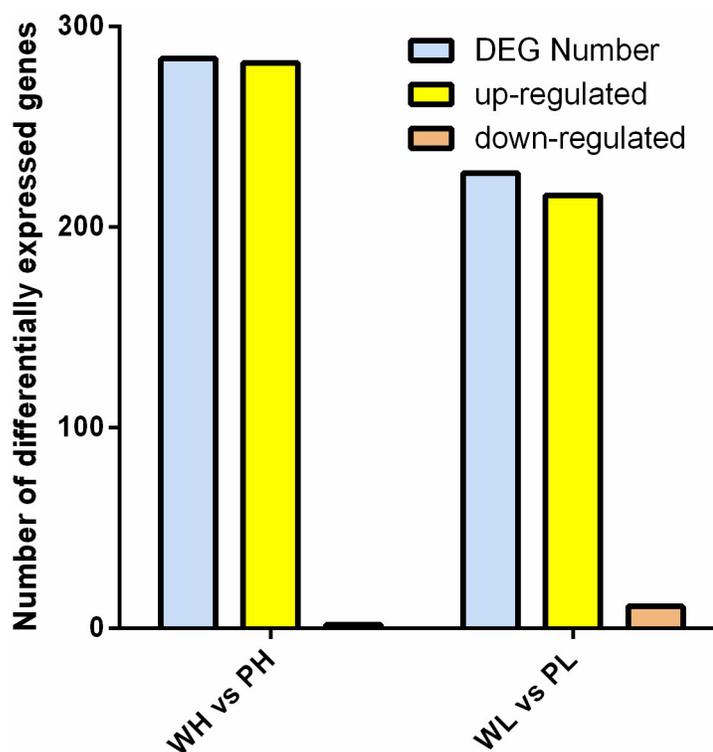


Figure 3 Analysis of DETs. Histogram of DETs resulted from comparisons of transcriptome data of WH vs. PH and WL vs. PL.

Full-size DOI: 10.7717/peerj.13688/fig-3

were classified as lincRNA (the long non-coding RNA of intergenic region), and only 16 (3.3%) lincRNAs were sorted into antisense-lincRNA (Fig. 2E).

Comparative analysis of differentially expressed transcripts (DET) profiling

DETs were identified between two cultivars. Data comparison of the groups PH vs. WH and WL vs. PL identified 284 and 227 DETs, respectively. Comparing the data between PH and WH, 282 was up-regulated, with only two genes down-regulated. With respect to the group PL vs. WL, 216 and 11 genes were up-/down-regulated, respectively (Fig. 3).

GO enrichment analysis found that the DETs were significantly enriched in biological process, molecular function, and cellular component. Within biological processes, DETs were mainly involved in translation and negative regulation of endopeptidase activity. Regarding molecular function, DETs were enriched in structural constituent of ribosome. As for cellular component, DETs were mostly enriched in ribosome and cytosolic large ribosomal subunit (Figs. 4A and 4B). KEGG enrichment analysis revealed DETs between PH and WH participated in the following pathways: flavonoid biosynthesis, glutathione metabolism, flavone and flavonol biosynthesis, and phenylpropanoid biosynthesis (Fig. 4C). In terms of DEGs between PL and WL, the related significant pathways were flavonoid biosynthesis pathway, phenylpropanoid biosynthesis pathway and glutathione metabolism pathway (Fig. 4D).

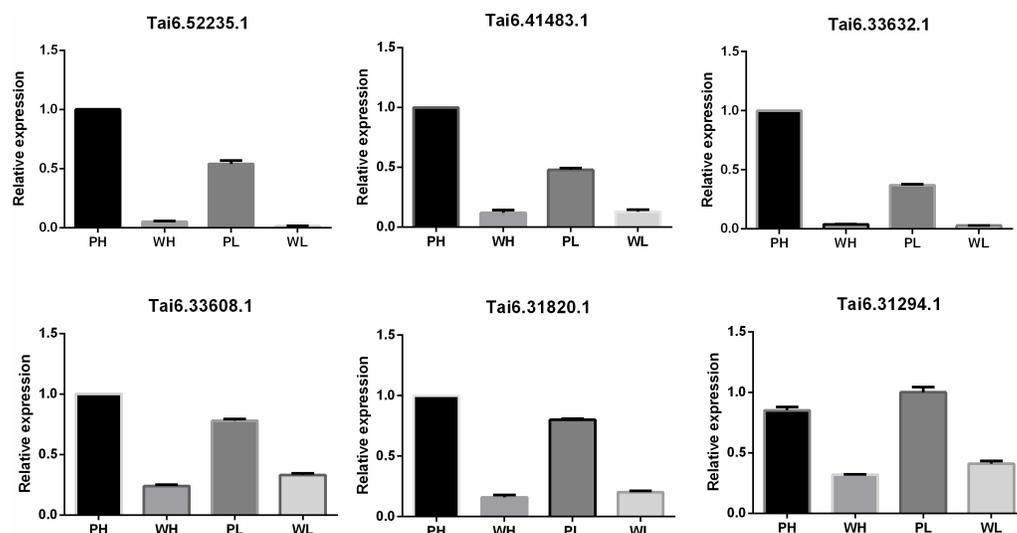


Figure 5 Quantitative real time PCR validation for six DETs participating in the flavonoid biosynthesis pathway.

Full-size DOI: [10.7717/peerj.13688/fig-5](https://doi.org/10.7717/peerj.13688/fig-5)

Validation of gene expression related to flavonoid biosynthesis pathway

To evaluate the accuracy and reproducibility of RNA-Seq data, DETs participating in the flavonoid biosynthesis pathway were selected for qRT-PCR verification. The primers used for qRT-PCR were displayed in [Table S7](#). Pfam annotation indicated Tai6.31294, Tai6.33608, and Tai6.31820 were DETs representing cytochrome P450. Tai6.41483 and Tai6.33632 belong to the genes encoding chalcone-flavanone isomerase. Tai6.52235 was the gene related to 2OG-Fe(II) oxygenase superfamily. The qRT-PCR confirmed that the expression tendency of these 6 transcripts was highly consistent with the RNA-Seq results ([Fig. 5](#)). Thus, the transcriptome data used for DET analysis were reliable.

DISCUSSION

Sweet potato is a special kind of anthocyanin-rich plant which benefits people' health worldwide ([Bakuradze et al., 2019](#)). Purple pigmentation of storage roots as one of the commonly evaluated characteristics of sweet potato varieties often arises attributed to the increase in the content and composition of anthocyanins ([Yoshimoto et al., 2001](#)). The researches indicated the content and composition of anthocyanins are distinct among different sweet potato varieties ([He et al., 2020](#)). In this work, we systematically investigated the variation in anthocyanin content among the storage roots of the two sweet potato cultivars viz. PFSP and WFSP at different developmental stages. According to the results of spectrophotometric measurement, the anthocyanin content in the root of PFSP was evidently much higher than that of WFSP at different developmental stages, suggesting that anthocyanin content variation gives rise to different pigmentation phenotypes of sweet potatoes.

In the present study, ONT RNA-Seq as one of the newest single-molecule long-read sequencing technologies was used to conduct full-length transcriptome sequencing of 12 samples obtained from tuberous roots of PFSP and WFSP. The clean data of each sample reached 2.75 GB, and the length of N50 ranged from 1,192 to 1,395 bp, which was much longer than those in previous studies using short read RNA-Seq (Huang *et al.*, 2012). It is well known that alternative splicing can select and utilize alternative splice sites in the pre-mRNA through different splicing events to generate multiple mRNA transcripts from a single gene (Feng *et al.*, 2019). Splicing events are involved in many physiological processes and play an important role in growth and development of plants (Ule & Blencowe, 2019), ONT RNA-Seq is conducive to finding different transcript isoforms of genes and providing strong evidence for alternative splicing identification. Here, based on the transcriptome data generated by ONT, we identified five main types of alternative splicing events including intron retention, alternative 3' splice site, alternative 5' splice site and exon skipping, which was in line with findings of previous reports on other plant species (Bush *et al.*, 2017).

Inferring lncRNA functions play an important role in elucidating the roles of lncRNAs in plant growth and adaptation, including the plant host–pathogen interaction. lncRNAs can form extremely complex secondary structures, and has multi-loop-stem structures, which can elaborate pleiotropy of lncRNAs. An increasing number of researches demonstrated that secondary structures of lncRNAs were more conservative, which is essential for lncRNA functions (Kim & Sung, 2012). In our research, a total of 486 lncRNAs were predicted, of which 343 (70.6%) lncRNAs were classified as lincRNA, and only 16 (3.3%) lncRNAs were sorted into antisense-lncRNA.

Subsequently, DETs profiling was conducted based on the transcriptome data, indicating most of DETs were up-regulated in PFSP as compared with WFSP. KEGG enrichment analysis revealed many DETs associated with flavonoid biosynthesis pathway including *CCoAOMT*, *PpLDOX*, *DFR*, Cytochrome P450, *CHI*, *CHS*, and 2OG-Fe(II) oxygenase superfamily were significantly up-regulated in PFSP cultivars. The up-regulated DETs might execute their significant regulatory functions for anthocyanin biosynthesis in PFSP cultivars. Accumulating evidence suggests the biological processes catalysed by *CHS*, *CHI*, *F3H*, *F3H ly*, *DFR*, *ANS* and 3 *GT* bring out the production of different anthocyanin subgroups by modifying the molecular skeleton and/or backbone. It was reported that a much higher transcript abundance of most of the structural genes including *IbC4H*, *IbCHS*, *IbCHI*, *IbF3H*, *IbDFR*, *IbANS*, and *IbUGT* was observed in PFSP cultivars when compared to WFSP and yellow- or orange-fleshed sweet potato (Wang *et al.*, 2018). Similarly, most of the structural genes involved in the anthocyanin biosynthesis pathway were up-regulated at fruit developmental stages of red pear as compared to its green color mutations (Yang *et al.*, 2015). Coordinated changes in expression of *DFR*, *CHI*, and *CHS* have also been found in differently colored Chinese bayberries (Niu *et al.*, 2010), grapes (Boss, Davies & Robinson, 1996), Arabidopsis (Saito *et al.*, 2013) and other plants (Wang *et al.*, 2017b). In the present study, comparative full-length transcriptome analysis by ONT revealed expression variation of important genes such as *CHS* and *DFR*, which significantly affected anthocyanin accumulation in storage roots of PFSP and WFSP. It could be inferred that transcriptional regulation of the flavonoid biosynthesis pathway based on transcriptome

data by ONT is beneficial to the accumulation of anthocyanin and other valuable flavonoids in sweet potatoes, thus improving economically agronomic traits and accelerating the breeding process.

CONCLUSIONS

Comparative full-length transcriptome analysis based on ONT serves as an effective approach to investigate the differences in anthocyanin accumulation in the storage roots of different sweet potato cultivars at transcript level, with the notable outcome that several key genes can now be closely linked to flavonoids biosynthesis. This study helps to improve understanding of molecular mechanism for anthocyanin accumulation in sweet potatoes and also provides a theoretical basis for high-quality sweet potato breeding.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was funded by the Fundamental Research Fund of Guangxi Academy of Agriculture Sciences (2021YT060) and the Guangxi Key Research and Development Plan Project (GuiKe AB16380085; GuiKe AB18221101). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Fundamental Research Fund of Guangxi Academy of Agriculture Sciences: 2021YT060.
Guangxi Key Research and Development Plan Project: AB16380085, AB18221101.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Jun Xiong performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Xiuhua Tang analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Minzheng Wei conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Wenjin Yu conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The sequencing data are available at the Sequence Read Archive (SRA), National Center for Biotechnology Information (NCBI): [PRJNA717378](https://www.ncbi.nlm.nih.gov/sra/PRJNA717378).

Data Availability

The following information was supplied regarding data availability:

The raw data and code are available in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.13688#supplemental-information>.

REFERENCES

- Alipour B, Rashidkhani B, Edalati S. 2016.** Dietary flavonoid intake, total antioxidant capacity and lipid oxidative damage: a cross-sectional study of Iranian women. *Nutrition* **32**:566–572 DOI [10.1016/j.nut.2015.11.011](https://doi.org/10.1016/j.nut.2015.11.011).
- Andreason SA, Olaniyi OG. 2021.** Large-scale seedling grow-out experiments do not support seed transmission of sweet potato leaf curl virus in sweet potato. *Plants* **10**(1):139–145 DOI [10.3390/plants10010139](https://doi.org/10.3390/plants10010139).
- Bakuradze T, Tausend A, Galan J, Groh IAM, Berry D. 2019.** Antioxidative activity and health benefits of anthocyanin-rich fruit juice in healthy volunteers. *Free Radical Research* **53**:1045–1055 DOI [10.1080/10715762.2019.1618851](https://doi.org/10.1080/10715762.2019.1618851).
- Benjamini Y, Hochberg Y. 1995.** Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57**:289–300.
- Boss PK, Davies C, Robinson SP. 1996.** Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L. cv shiraz grape berries and the implications for pathway regulation. *Plant Physiology* **111**:1059–1066 DOI [10.1104/pp.111.4.1059](https://doi.org/10.1104/pp.111.4.1059).
- Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO. 2017.** Alternative splicing and the evolution of phenotypic novelty. *Philosophical Transactions of the Royal Society B* **372**(1713):20150474 DOI [10.1098/rstb.2015.0474](https://doi.org/10.1098/rstb.2015.0474).
- De Moura FF, Miloff A, Boy E. 2015.** Retention of provitamin a carotenoids in staple crops targeted for biofortification in Africa: cassava, maize and sweet potato. *Critical Reviews in Food Science and Nutrition* **55**:1246–1269 DOI [10.1080/10408398.2012.724477](https://doi.org/10.1080/10408398.2012.724477).
- Feng X, Jia Y, Zhu R, Chen K, Chen Y. 2019.** Characterization and analysis of the transcriptome in *Gymnocypripis selincuoensis* on the Qinghai-Tibetan Plateau using single-molecule long-read sequencing and RNA-seq. *DNA Research* **26**:353–363 DOI [10.1093/dnares/dsz014](https://doi.org/10.1093/dnares/dsz014).
- Ghasemzadeh A, Talei D, Jaafar HZ, Juraimi AS, Mohamed MT, Puteh A, Halim MR. 2016.** Plant-growth regulators alter phytochemical constituents and pharmaceutical quality in Sweet potato (*Ipomoea batatas* L.). *BMC Complementary and Alternative Medicine* **16**:152 DOI [10.1186/s12906-016-1113-1](https://doi.org/10.1186/s12906-016-1113-1).
- He L, Liu X, Liu S, Zhang J, Zhang Y, Sun Y, Tang R, Wang W, Cui H, Li R, Zhu H, Jia X. 2020.** Transcriptomic and targeted metabolomic analysis identifies genes

- and metabolites involved in anthocyanin accumulation in tuberous roots of sweetpotato (*Ipomoea batatas* L.). *Plant Physiology and Biochemistry* **156**:323–332 DOI [10.1016/j.plaphy.2020.09.021](https://doi.org/10.1016/j.plaphy.2020.09.021).
- Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V. 2011.** Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *Journal of Experimental Botany* **62**:2465–2483 DOI [10.1093/jxb/erq442](https://doi.org/10.1093/jxb/erq442).
- Hoang PNT, Michael TP, Gilbert S, Chu P, Motley ST, Appenroth KJ, Schubert I. 2018.** Generating a high-confidence reference genome map of the Greater Duckweed by integration of cytogenomic, optical mapping, and Oxford Nanopore technologies. *The Plant Journal* **96**:670–684 DOI [10.1111/tpj.14049](https://doi.org/10.1111/tpj.14049).
- Hong HT, Netzel ME, O'Hare TJ. 2020.** Optimisation of extraction procedure and development of LC-DAD-MS methodology for anthocyanin analysis in anthocyanin-pigmented corn kernels. *Food Chemistry* **319**:126515–126522 DOI [10.1016/j.foodchem.2020.126515](https://doi.org/10.1016/j.foodchem.2020.126515).
- Hu Z, Zhang Y, He Y, Cao Q, Zhang T, Lou L, Cai Q. 2020.** Full-length transcriptome assembly of italian ryegrass root integrated with RNA-Seq to identify genes in response to plant cadmium stress. *International Journal of Molecular Sciences* **21**(3):1067 DOI [10.3390/ijms21031067](https://doi.org/10.3390/ijms21031067).
- Huang L, Dong H, Zhou D, Li M, Liu Y, Zhang F, Feng Y, Yu D, Lin S, Cao J. 2018.** Systematic identification of long non-coding RNAs during pollen development and fertilization in *Brassica rapa*. *The Plant Journal* **96**:203–222 DOI [10.1111/tpj.14016](https://doi.org/10.1111/tpj.14016).
- Huang HH, Xu LL, Tong ZK, Lin EP, Liu QP, Cheng LJ, Zhu MY. 2012.** *De novo* characterization of the Chinese fir (*Cunninghamia lanceolata*) transcriptome and analysis of candidate genes involved in cellulose and lignin biosynthesis. *BMC Genomics* **13**:648 DOI [10.1186/1471-2164-13-648](https://doi.org/10.1186/1471-2164-13-648).
- Jaakola L. 2013.** New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends in Plant Science* **18**:477–483 DOI [10.1016/j.tplants.2013.06.003](https://doi.org/10.1016/j.tplants.2013.06.003).
- Kim ED, Sung S. 2012.** Long noncoding RNA: unveiling hidden layer of gene regulatory networks. *Trends in Plant Science* **17**:16–21 DOI [10.1016/j.tplants.2011.10.008](https://doi.org/10.1016/j.tplants.2011.10.008).
- Kovinich N, Kayanja G, Chanoca A, Riedl K, Otegui MS, Grotewold E. 2014.** Not all anthocyanins are born equal: distinct patterns induced by stress in *Arabidopsis*. *Planta* **240**:931–940 DOI [10.1007/s00425-014-2079-1](https://doi.org/10.1007/s00425-014-2079-1).
- Li Y, Dai C, Hu C, Liu Z, Kang C. 2017.** Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry. *The Plant Journal* **90**:164–176 DOI [10.1111/tpj.13462](https://doi.org/10.1111/tpj.13462).
- Li Q, Kou M, Li C, Zhang YG. 2021.** Comparative transcriptome analysis reveals candidate genes involved in anthocyanin biosynthesis in sweetpotato (*Ipomoea batatas* L.). *Plant Physiology and Biochemistry* **158**:508–517 DOI [10.1016/j.plaphy.2020.11.035](https://doi.org/10.1016/j.plaphy.2020.11.035).
- Lin KT, Krainer AR. 2019.** PSI-Sigma: a comprehensive splicing-detection method for short-read and long-read RNA-seq analysis. *Bioinformatics* **35**:5048–5054 DOI [10.1093/bioinformatics/btz438](https://doi.org/10.1093/bioinformatics/btz438).

- Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2\Delta\Delta CT$ method. *Methods* 25:402–408 DOI 10.1006/meth.2001.1262.
- Man YP, Wang YC, Li ZZ, Jiang ZW, Yang HL, Gong JJ, He SS, Wu SQ, Yang ZQ, Zheng J, Wang ZY. 2015. High-temperature inhibition of biosynthesis and transportation of anthocyanins results in the poor red coloration in red-fleshed *Actinidia chinensis*. *Physiologia Plantarum* 153:565–583 DOI 10.1111/ppl.12263.
- Niu SS, Xu CJ, Zhang WS, Zhang B, Li X, Lin-Wang K, Ferguson IB, Allan AC, Chen KS. 2010. Coordinated regulation of anthocyanin biosynthesis in Chinese bayberry (*Myrica rubra*) fruit by a R2R3 MYB transcription factor. *Planta* 231:887–899 DOI 10.1007/s00425-009-1095-z.
- Petroni K, Tonelli C. 2011. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Science* 181:219–229 DOI 10.1016/j.plantsci.2011.05.009.
- Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, Fernie AR. 2013. The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. *Plant Physiology and Biochemistry* 72:21–34 DOI 10.1016/j.plaphy.2013.02.001.
- Slatko BE, Gardner AF, Ausubel FM. 2018. Overview of next-generation sequencing technologies. *Current Protocols in Molecular Biology* 122:e59 DOI 10.1002/cpmb.59.
- Sun H, Mu T, Xi L, Zhang M, Chen J. 2014. Sweet potato (*Ipomoea batatas* L.) leaves as nutritional and functional foods. *Food Chemistry* 156:380–389 DOI 10.1016/j.foodchem.2014.01.079.
- Sun Y, Pan Z, Yang C, Jia Z, Guo X. 2019. Comparative assessment of phenolic profiles, cellular antioxidant and antiproliferative activities in ten varieties of sweet potato (*Ipomoea Batatas*) storage roots. *Molecules* 24(24):4476–4484 DOI 10.3390/molecules24244476.
- Uapinyoying P, Goecks J. 2020. A long-read RNA-seq approach to identify novel transcripts of very large genes. *Genome Research* 30:885–897 DOI 10.1101/gr.259903.119.
- Ule J, Blencowe BJ. 2019. Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Molecular Cell* 76:329–345 DOI 10.1016/j.molcel.2019.09.017.
- Wang Z, Cui Y, Vainstein A, Chen S, Ma H. 2017b. Regulation of fig (*Ficus carica* L.) fruit color: metabolomic and transcriptomic analyses of the flavonoid biosynthetic pathway. *Frontiers in Plant Science* 8:1990 DOI 10.3389/fpls.2017.01990.
- Wang A, Li R, Ren L, Gao X, Zhang Y, Ma Z, Ma D, Luo Y. 2018. A comparative metabolomics study of flavonoids in sweet potato with different flesh colors (*Ipomoea batatas* (L.) Lam). *Food Chemistry* 260:124–134 DOI 10.1016/j.foodchem.2018.03.125.
- Wang T, Wang H, Cai D, Gao Y, Zhang H, Wang Y, Lin C, Ma L, Gu L. 2017a. Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*). *The Plant Journal* 91:684–699 DOI 10.1111/tpj.13597.

- Wang H, Yang Q, Ferdinand U, Gong X, Qu Y, Gao W, Ivanistau A, Feng B, Liu M. 2020.** Isolation and characterization of starch from light yellow, orange, and purple sweet potatoes. *International Journal of Biological Macromolecules* **160**:660–668 DOI [10.1016/j.ijbiomac.2020.05.259](https://doi.org/10.1016/j.ijbiomac.2020.05.259).
- Yang Y, Yao G, Yue W, Zhang S, Wu J. 2015.** Transcriptome profiling reveals differential gene expression in proanthocyanidin biosynthesis associated with red/green skin color mutant of pear (*Pyrus communis* L.). *Frontiers in Plant Science* **6**:795 DOI [10.3389/fpls.2015.00795](https://doi.org/10.3389/fpls.2015.00795).
- Yao S, Liang F, Gill RA, Huang J, Cheng X, Liu Y, Tong C, Liu S. 2020.** A global survey of the transcriptome of allopolyploid *Brassica napus* based on single-molecule long-read isoform sequencing and Illumina-based RNA sequencing data. *The Plant Journal* **103**:843–857 DOI [10.1111/tpj.14754](https://doi.org/10.1111/tpj.14754).
- Yoshimoto M, Okuno S, Yamaguchi M, Yamakawa O. 2001.** Antimutagenicity of deacylated anthocyanins in purple-fleshed sweetpotato. *Bioscience, Biotechnology, and Biochemistry* **65**:1652–1655 DOI [10.1271/bbb.65.1652](https://doi.org/10.1271/bbb.65.1652).
- Younesi-Melerdi E, Nematzadeh GA, Pakdin-Parizi A. 2020.** *De novo* RNA sequencing analysis of *Aeluropus littoralis* halophyte plant under salinity stress. *Scientific Reports* **10**:9148 DOI [10.1038/s41598-020-65947-5](https://doi.org/10.1038/s41598-020-65947-5).
- Zhang G, Sun M, Wang J, Lei M, Li C, Zhao D, Huang J, Li W, Li S, Li J, Yang J, Luo Y, Hu S, Zhang B. 2019.** PacBio full-length cDNA sequencing integrated with RNA-seq reads drastically improves the discovery of splicing transcripts in rice. *The Plant Journal* **97**:296–305 DOI [10.1111/tpj.14120](https://doi.org/10.1111/tpj.14120).
- Zhao L, Zhang H, Kohnen MV, Prasad K, Gu L, Reddy ASN. 2019.** Analysis of transcriptome and epitranscriptome in plants using PacBio iso-seq and nanopore-based direct RNA sequencing. *Frontiers in Genetics* **10**:253 DOI [10.3389/fgene.2019.00253](https://doi.org/10.3389/fgene.2019.00253).